

# Mean Entropies

B. H. Lavenda\*

Università degli Studi, Camerino 62032 (MC) Italy

(Dated: February 2, 2008)

Entropies must correspond to mean values for them to be measurable. The Shannon entropy corresponds to the weighted arithmetic mean, whereas the Rényi entropy corresponds to the exponential mean. These means refer to code lengths, which are converted into entropies by replacing the length of a sequence by the negative logarithm of the probability of its occurrence. Only affine and exponential generating functions of means preserve the property of additivity and invariance under translations, and hence are Kolmogorov-Nagumo functions, resulting in the Shannon and Rényi entropies, respectively. Pseudo-additive entropies are generating functions of means of order  $0 \leq \tau < 1$ , which is the exponential of the Rényi entropy, or in the  $\tau = 0$  limit, the Shannon entropy. Means of any order cannot be expressed as escort averages because such averages contradict the fact that the means are monotonically increasing functions of their order. Exponential mean error functions of Rényi, in general, and Shannon, in particular, are shown to be measures of the extent of a distribution.

PACS numbers: 05.70.Ln, 05.20.Gg, 05.40.-a

## ENTROPY AND CODING THEORY

In his seminal paper on entropy and information, Rényi [1] laid down the fundamental properties of entropy and its relation to a measure of information. Many of these tenets have been all but abandoned [2].

Basing himself on the relation between the mean code length,

$$L_a(n) = \sum_{i=1}^N p_i n_i \quad (1)$$

and the Shannon entropy, Rényi argued convincingly that any putative candidate for an entropy should be a mean. Here,  $p = (p(x_1), p(x_2), \dots, p(x_N))$  is the set of probabilities of  $N$  input symbols,  $x = (x_1, x_2, \dots, x_N)$ , that are to be encoded using an alphabet of size  $D$ . Each  $p(x_i) > 0$ , and the distribution is complete,  $\sum_{i=1}^N p(x_i) = 1$ . The  $x_i$  represent a sequence of  $n_i$  characters taken from the alphabet.

There exists a uniquely decipherable code with lengths  $n_i$  iff

$$\sum_{i=1}^N D^{-n_i} \leq 1, \quad (2)$$

which is known as Kraft's inequality [3]. The equality is automatically guaranteed by setting

$$n_i = -\log_D p(x_i), \quad (3)$$

which represents the amount of information received by knowing that an event of probability  $p(x_i)$  has occurred.

Introducing (3) into the weighted arithmetic average (1) gives Shannon's expression for the entropy,

$$S_S(p) = -\sum_{i=1}^N p_i \log_D p_i, \quad (4)$$

where the abbreviation  $p_i = p(x_i)$  has been introduced. Rényi addressed the question of what other entropies are obtainable when the weighted arithmetic mean, (4), is replaced by a generalized mean [4]

$$S(n) = \phi^{-1} \left\{ \sum_{i=1}^N p_i \phi(n_i) \right\} = \mathfrak{M}_\phi(n), \quad (5)$$

where  $\phi(x)$  is a strictly monotonic and continuous function that possesses an inverse  $\phi^{-1}(x)$ .

The generating function  $\phi$  must be so chosen that the generalized mean (5) possess the following properties. First and foremost, it must have the property of *additivity*. If  $q = (q(x_1), q(x_2), \dots, q(x_N))$  represents another finite discrete probability distribution then the entropy of their direct product should be additive

$$S(p \otimes q) = S(p) + S(q). \quad (6)$$

This is guaranteed by the fact that the events are independent so that their probabilities multiply, and the entropy satisfies the functional equation, (6). According to (3) multiplicative probabilities means that code lengths are additive, as they should be.

Secondly, the entropy should possess the *mean-value property*, which says that the entropy of the union of two distributions is the weighted arithmetic mean of the individual entropies.

Other properties listed by Rényi [1] that an entropy should have were symmetry, continuity, and normalization.

---

\*Electronic address: bernard.lavenda@unicam.it

## EQUIVALENT MEANS

In addition to these properties, any *classical* expression for the entropy should be *translation invariant*, which says that only entropy differences are measurable. For  $N = 2$ , the generalized mean must satisfy

$$\begin{aligned} \phi^{-1} [p_1 \phi(x_1 + a) + p_2 \phi(x_2 + a)] \\ = \phi^{-1} [p_1 \phi(x_1) + p_2 \phi(x_2)] + a, \end{aligned} \quad (7)$$

where  $a$  is a constant. Such functions,  $\phi$ , were first investigated by Kolmogorov [5] and Nagumo [6], and will be referred to as KN functions. The only known strictly monotonic increasing solutions of the functional equation (7) are [7]

$$\phi(x) = ax + b$$

affine, and

$$\phi(x) = aD^{\tau x} + b,$$

exponential functions, where  $a \neq 0$  and  $\tau \neq 0$ .

The affine solution leads immediately to the Shannon entropy (4) under condition (3), whereas the exponential solution leads one to consider exponential mean lengths [8]

$$L_e(n) = \frac{1}{\tau} \log_D \sum_{i=1}^N p_i D^{\tau n_i} \quad (8)$$

whose corresponding entropies are [1]

$$S_R(p) = \frac{1}{\tau} \log_D \sum_{i=1}^N p_i^{1-\tau} \quad (9)$$

under condition (3).

It is imperative to emphasize that Rényi worked with the code lengths as the independent variables, and not with the number of different sequences of length  $n_i$  [9]:

$$\omega_i = D^{n_i}. \quad (10)$$

For if he had considered (10) as the set of independent variables he would have obtained means of order  $\tau$ ,

$$\mathfrak{M}_\tau(\omega) = \left( \sum_{i=1}^N p_i \omega_i^\tau \right)^{1/\tau},$$

leading to the expression

$$S(p) = \left( \sum_{i=1}^N p_i^{1-\tau} \right)^{1/\tau} \quad (11)$$

as the entropy, rather than as the ‘exponential entropy’ [10]. Such an entropy would not possess the property of additivity that its logarithm would restore.

## PSEUDO-ADDITIVITY

The foregoing discussions provides a basis for understanding the property of ‘pseudo-additivity’ that certain entropies have been found to possess [11, 12].

Consider the function

$$\phi(n_i) = \frac{D^{\tau n_i} - 1}{\tau} \quad (12)$$

which becomes affine in the limit  $\tau \rightarrow 0$ . Since (12) is a KN function, its mean is equivalent to (8), and, consequently, to the Rényi entropy, (9), when condition (3) is imposed. However, were we to introduce (3) directly into (12),

$$\phi(1/p_i) = \frac{p_i^{-\tau} - 1}{\tau} \quad (13)$$

and then take its mean value, we would obtain an exponential entropy,

$$\phi^{-1} \left\{ \sum_{i=1}^N p_i \phi(1/p_i) \right\} = \mathfrak{M}_\tau(1/p) = D^{S_R(p)},$$

rather than the Rényi entropy itself.

The weighted arithmetic mean of (13),

$$S_{HC}(p) = \sum_{i=1}^N p_i \phi(1/p_i) = \frac{\sum_{i=1}^N p_i^{1-\tau} - 1}{\tau}, \quad (14)$$

is known as the Havrda-Charvát [13]-Daróczy [14]-Tsallis [2] (HCDT) entropy. The HCDT entropy, (14), is pseudo-additive in that it is the solution of

$$S_{HC}(p \otimes q) = S(p) + S(q) + \tau S(p) S(q), \quad (15)$$

and not of (6). The KN function, whose mean is equivalent to the mean of (13),

$$\mathfrak{M}_{\tilde{\phi}}(1/p) = \mathfrak{M}_\phi(1/p),$$

is

$$\tilde{\phi}(1/p_i) = \tau \phi(1/p_i) + 1.$$

Its weighted arithmetic mean,

$$\tilde{S}_{HC}(p) = \tau S_{HC}(p) + 1,$$

satisfies the functional equation

$$\tilde{S}_{HC}(p \otimes q) = \tilde{S}_{HC}(p) \tilde{S}_{HC}(q), \quad (16)$$

implying a power law solution [7, p. 39], and the complete loss of additivity. Hence, on the basis of equivalent means we can transform pseudo-additivity, (15), into a multiplicative relation, (16), showing that neither relation has any thermodynamic meaning regarding the lack of extensivity.

As (11) clearly shows, it is the logarithm of the mean of the HCDT entropy that has physical meaning, and this is the Rényi entropy. If we had insisted on working with code lengths, and not with their probabilities, we would have obtained the Rényi entropy directly from introducing (3) into the mean code length (8).

## EXPONENTIAL MEAN ENTROPY BOUNDS ON MEAN NUMBER OF SEQUENCES

According to Jensen's inequality for a convex function

$$\sum_{i=1}^N p_i D^{\tau n_i} \geq D^{\tau \sum_{i=1}^N p_i n_i}. \quad (17)$$

Taking logarithms and considering  $\tau > 0$ , there results

$$\frac{1}{\tau} \log_D \left( \sum_{i=1}^N p_i D^{\tau n_i} \right) \geq \sum_{i=1}^N p_i n_i, \quad (18)$$

asserting that the exponential mean of parameter  $\tau > 0$  is never inferior to the weighted arithmetic mean. This is most easily shown for small  $\tau$ . Expanding the exponent and then the logarithm in powers of  $\tau$ , we get to lowest order [15]

$$\begin{aligned} & \frac{1}{\tau} \log_D \sum_{i=1}^N p_i D^{\tau n_i} \\ &= \sum_i p_i n_i + \frac{\tau}{2} \left\{ \sum_{i=1}^N p_i n_i^2 - \left( \sum_{i=1}^N p_i n_i \right)^2 \right\} + \dots \end{aligned}$$

The term in the curly brackets is always positive since it is the variance of  $n_i$ . We must also restrict  $\tau < 1$  in order to ensure that the Rényi entropy is concave.

On the strength of (3), inequality (18) shows that the Rényi entropy is bounded from below by the Shannon entropy, and since the entropy is maximum for a uniform distribution, we have the following hierarchy:

$$S_S \leq S_R \leq S_H,$$

where  $S_H = \log_D N$  is the Hartley entropy.

In terms of the number of different sequences of length  $n_i$ , Jensen's inequality (17) becomes

$$\mathfrak{M}_\tau(\omega) = \left( \sum_{i=1}^N p_i \omega_i^\tau \right)^{1/\tau} \geq \prod_{i=1}^N \omega_i^{p_i} = \mathfrak{M}_0(\omega),$$

which says that means of order  $\tau > 0$  can never be inferior to the geometric mean [16], the weighted arithmetic mean ( $\tau = 1$ ) being a particular case. The mean inequality for the same order and different argument [4, p. 14],

$$\mathfrak{M}_\tau(\omega) \geq \mathfrak{M}_\tau(1/p), \quad (19)$$

follows from the fact that according to the Kraft inequality  $\omega_i \geq p_i^{-1}$  for all  $i$ . The equality in (19) holds when (3) is satisfied. The mean number of sequences is bounded from below by the exponential of the Rényi entropy. As a problem in majorization, we say that  $\omega$  majorizes  $p^{-1}$ ,  $\omega \succ p^{-1}$ .

A similar, but not identical, result was found by Campbell [8, 17], who used Hölder's inequality [4, p. 24]

$$\mathfrak{M}_{\tau/\alpha}(\omega) \mathfrak{M}_1(1/\omega p) \geq \mathfrak{M}_\tau(1/p),$$

where  $\alpha = 1 - \tau$ , and the Kraft inequality, (2), to obtain

$$\mathfrak{M}_{\tau/\alpha}(\omega) \geq \mathfrak{M}_\tau(1/p). \quad (20)$$

The condition for the equality in (20) is *not* (3), but, rather,

$$\omega_i^{-1} = p_i^\alpha / \sum_{i=1}^N p_i^\alpha, \quad (21)$$

which has been referred to as an 'escort probability' [18]. Whereas (19) implies (20), the converse is not true. In other words, inequality (20) holds for  $\alpha = 1$ , as (19) clearly shows, so that (21) must also hold for  $\alpha = 1$ , which is (3). In other words, (3) is sharper than (21). Moreover, since inequality (20) holds for  $\alpha < 1$ , the same must be true in (21). This condition has apparently gone unappreciated [19].

## 'ESCORT' AVERAGES

Escort averaging has been used in variational formulations that maximize the pseudo-additive entropy (14) with respect to escort expectations of thermodynamic constraints [19]. Pragmatically speaking, it leads to analytic expressions for the variational equations, which would otherwise not exist. If escort averaging has any meaning at all, it must yield viable expressions for the means, and functions of the means.

Mean entropies are special cases of generalized means, where the variables,  $-\log p_i$ , and their weights,  $p_i$ , are not independent. Rather than considering means of order  $\tau$ , for which a demonstration that the mean is a monotonically increasing function of its order is given in [15], we shall consider the exponential entropy (11) and the Rényi entropy (9).

Differentiating

$$S^\tau(p) = \sum_{i=1}^N p_i^{1-\tau},$$

with respect to  $\tau$  yields

$$S^\tau \left[ \log_D S + \tau \frac{d \log_D S}{d\tau} \right] = - \sum_{i=1}^N p_i^{1-\tau} \log_D p_i. \quad (22)$$

At a stationary point,  $d \log_D S / d\tau = 0$ , and

$$S_R = \log_D S = - \sum_{i=1}^N p_i^\alpha \log_D p_i / \sum_{i=1}^N p_i^\alpha, \quad (23)$$

where again  $\alpha = 1 - \tau$ . As  $\alpha \rightarrow 1$  the entropy (23) transforms into the Shannon entropy, (4). Expression (23) states that the logarithm of the exponential entropy (11), which is the Rényi entropy, is the escort average of  $-\log_D p_i$ . The second derivative of (22), evaluated at the stationary point is

$$\begin{aligned} & \tau \frac{d^2 \log_D S}{d\tau^2} \\ &= \frac{\sum_{i=1}^N p_i^\alpha (\log_D p_i)^2}{\sum_{i=1}^N p_i^\alpha} - \left( \frac{\sum_{i=1}^N p_i^\alpha \log_D p_i}{\sum_{i=1}^N p_i^\alpha} \right)^2 > 0, \end{aligned}$$

because the right-hand side is the variance of  $-\log_D p_i$  under escort averaging. Hence,  $d^2 \log_D S / d\tau^2 > 0$  for  $\tau > 0$ , and  $d^2 \log_D S / d\tau^2 < 0$  for  $\tau < 0$ , implying that there are two extrema: a local maximum for  $\tau < 0$  and a local minimum for  $\tau > 0$ . This requires  $d \log_D S / d\tau \leq 0$  at  $\tau = 0$ , where the equality sign applies to the degenerate case where the extremes coincide in a point of inflection at  $\tau = 0$ .

Now, if it can be shown that  $d \log_D S / d\tau > 0$  at  $\tau = 0$ , then  $\log_D S$  has no extrema as a function of  $\tau$ , and, is, in fact, a monotonically increasing function of  $\tau$  for all values of  $\tau$ . From this we will conclude that  $\log_D S$  cannot be expressed as an escort average (23), derived from a stationary condition, since such a condition does not exist.

Writing (22) in the form

$$S^\tau \frac{d \log_D S}{d\tau} = -\frac{1}{\tau} \left\{ \sum_{i=1}^N p_i^{1-\tau} \log_D p_i + S^\tau \log_D S \right\},$$

it is apparent that the ratio on the right-hand side is of the form  $0/0$  as  $\tau \rightarrow 0$  because  $\log_D S \rightarrow S_S$  in that limit. With the aid of L'Hôpital's rule we get

$$2 \lim_{\tau \rightarrow 0} \left( S^\tau \frac{d \log_D S}{d\tau} \right) = \sum_{i=1}^N p_i (\log_D p_i)^2 - S_S^2 > 0.$$

The inequality follows from the fact that the right-hand side is the variance of  $-\log_D p_i$ . Hence,  $d \log_D S / d\tau > 0$  as  $\tau \rightarrow 0$ , and so it is positive for all  $\tau$ . This implies that the logarithm of the exponential entropy (11) is an increasing function of  $\tau$ , and that no stationary point given by condition (23) exists.

This is easily confirmed from the expression for the Rényi entropy. Differentiating (9) with respect to  $\tau$  gives

$$\frac{dS_R}{d\tau} = -\frac{1}{\tau} \left[ S_R + \frac{\sum_{i=1}^N p_i^{1-\tau} \log_D p_i}{\sum_{i=1}^N p_i^{1-\tau}} \right].$$

The stationary condition is again given by the escort average, (23). The product  $\tau d^2 S_R / d\tau^2 > 0$  at the stationary point so that we have a local minimum for  $\tau > 0$  and a local maximum for  $\tau < 0$ . This means that the curve of  $S_R$  versus  $\tau$  has a negative slope as it passes through

$\tau = 0$ . The demonstration that  $S_R$  has no extrema when considered as a function of  $\tau$  follows exactly as before.

Hence, the Rényi entropy, or for that matter any mean [15], cannot be expressed as an escort average because that would violate the condition that the mean is an increasing function of  $\tau$ . It is this property, in fact, which guarantees that the arithmetic mean ( $\tau = 1$ )  $\geq$  geometric mean ( $\tau = 0$ )  $\geq$  harmonic mean ( $\tau = -1$ ).

## EXTENT OF A DISTRIBUTION

In the next to the last section we have found that the exponential of the Rényi entropy is the lower bound on the mean value of the number of different sequences of order  $\tau$ , whose equivalent mean was the mean of the HCDT entropy. Exponential mean entropies have been shown to be measures of the extent of a distribution [10].

The measure of the extent of a distribution is inherently related to the error that is committed by using an estimated probability distribution,  $q$ , when the 'true' probability distribution is  $p$ . Rényi [20] has referred to this as 'information gain', while Kullback [21] used the term 'directed divergence', being based on the Shannon inequality

$$\mathcal{E}_S(q|p) = \sum_{i=1}^N p_i \log_D \left( \frac{p_i}{q_i} \right) \geq 0. \quad (24)$$

The quantity  $-\sum_{i=1}^N p_i \log_D q_i$  is referred to as the inaccuracy [22]. Inequality (24) is easily seen to be a consequence of the arithmetic-geometric mean inequality:

$$\prod_{i=1}^N \left( \frac{q_i}{p_i} \right)^{p_i} \leq \sum_{i=1}^N p_i \left( \frac{q_i}{p_i} \right) = 1. \quad (25)$$

As a generalization of (24) we may consider the generating function

$$\phi(q_i/p_i) = \frac{(q_i/p_i)^\tau - 1}{\tau} \quad (26)$$

since the negative of its weighted arithmetic average in the limit as  $\tau \rightarrow 0$  is

$$-\lim_{\tau \rightarrow 0} \sum_{i=1}^N p_i \phi(q_i/p_i) = \mathcal{E}_S(q|p).$$

The mean of (26) in the same limit is

$$\lim_{\tau \rightarrow 0} \phi^{-1} \left\{ \sum_{i=1}^N p_i \phi(q_i/p_i) \right\} = D^{-\mathcal{E}_S(q|p)} \leq 1,$$

which we will see to be related to the extent of a distribution.

The arithmetic average of (26),

$$\sum_{i=1}^N p_i \phi(q_i/p_i) = \frac{\sum_{i=1}^N p_i^{1-\tau} q_i^\tau - 1}{\tau},$$

has been referred to as the error incurred when the distribution  $q$  is used instead of the ‘true’ distribution,  $p$  [11, p. 208]. The mean value of (26) has the equivalent mean

$$\mathfrak{M}_\phi(q/p) = \mathfrak{M}_\tau(q/p) = \left( \sum_{i=1}^N p_i^{1-\tau} q_i^\tau \right)^{1/\tau} \leq 1, \quad (27)$$

since  $\phi(x)$  is a linear function of  $x^\tau$  for  $\tau \neq 0$  [4, p. 68]. The inequality in (27) follows directly from Hölder’s inequality,

$$\sum_{i=1}^N p_i^{1-\tau} q_i^\tau \leq \left( \sum_{i=1}^N p_i \right)^{1-\tau} \left( \sum_{i=1}^N q_i \right)^\tau \leq 1,$$

for  $0 < \tau < 1$ . The second inequality makes allowance for incomplete distributions.

The error function with a parameter  $\tau$ ,

$$\mathcal{E}_R(q|p) = -\frac{1}{\tau} \log_D \left( \sum_{i=1}^N p_i^{1-\tau} q_i^\tau \right) \geq 0, \quad (28)$$

is constructed in analogy to the Rényi entropy [11, p. 208]. In the  $\tau = 0$  limit, (28) reduces to the Shannon error function (24).

Because of the inequality of the means [4, p. 26]

$$\mathfrak{M}_r(x) \leq \mathfrak{M}_s(x),$$

for  $r < s$ , with equality iff the probability distribution is uniform, the mean

$$\mathfrak{M}_\tau(q/p) = D^{-\mathcal{E}_R(q|p)} \quad (29)$$

provides a measure of the extent of a distribution. The mean (29) is homogeneous in  $q$ , and satisfies

$$D^{-\mathcal{E}_S(q|p)} \leq D^{-\mathcal{E}_R(q|p)} \leq \sum_{i=1}^N q_i. \quad (30)$$

The lower limit corresponds to the geometric mean

$$\mathfrak{M}_0(q/p) = \prod_{i=1}^N \left( \frac{q_i}{p_i} \right)^{p_i} = D^{-\mathcal{E}_S(q|p)},$$

and is least affected by variations in  $q/p$  [10], while the upper limit

$$\mathfrak{M}_1(q) = \sum_{i=1}^N q_i$$

constitutes the ‘range’ of the distribution. The range is most elementary measure of the extent of a distribution.

The difference between an ordinary and an *incomplete* random variable is that the latter is not defined at every point in the sample space [20, p. 570]. Points where the random variable are undefined are said to be unobservable, and  $\sum_{i=1}^N q_i < 1$  is the probability that the outcome will be observable.

The arithmetic-geometric mean inequality (25) can be written as

$$\sum_{i=1}^N q_i \geq D^{\sum_{i=1}^N (p_i \log_D q_i - p_i \log_D p_i)}, \quad (31)$$

which is the first and second inequalities in (30). Shannon’s inequality (24) guarantees that the right-hand side of (31) is less than unity. The equality sign holds in (31) when  $p_i = \lambda q_i$ , where  $\lambda = \left( \sum_{i=1}^N q_i \right)^{-1}$ , thus ensuring that  $p$  is a complete distribution. This condition is obtained by requiring that  $-\sum_{i=1}^N p_i \log_D q_i$  be a minimum subject to  $\sum_{i=1}^N q_i = \text{const}$ . In fact, an essential part of the coding theorem for a noiseless channel is to show that the minimum of  $-\sum_{i=1}^N p_i \log q_i$  is  $-\sum_{i=1}^N p_i \log_D p_i$  subject to the constraint  $\sum_{i=1}^N q_i = 1$  [3, pp. 17–20]. The latter constraint implies  $\lambda = 1$ , and the Shannon error function vanishes leading to an equality in (31).

However, there is another way the equality can be satisfied in (31): The minimum of  $\sum_{i=1}^N q_i$  subject to the constraint  $\sum_{i=1}^N p_i \log_D q_i = \text{const}$  is  $D^{-\mathcal{E}_S(q|p)}$  [10]. Introducing the Lagrange multiplier  $\lambda = q_i/p_i$ , where  $\lambda = \sum_{i=1}^N q_i$ , into the Shannon error function (24) results in

$$\mathcal{E}_S(q|p) = \log_D \left( 1 / \sum_{i=1}^N q_i \right). \quad (32)$$

An incomplete distribution  $\sum_{i=1}^N q_i < 1$  leads to a finite error,  $\mathcal{E}_S > 0$ . The argument in the logarithm of (32) can be interpreted as the number of digits necessary to specify the set of observable events. The smaller the set, the greater the number of digits that will be required.

In this respect, (32) is the antithesis of Boltzmann’s principle, where  $\left( \sum_{i=1}^N q_i \right)^{-1}$  is paired to the ‘thermodynamic probability’, and the Shannon error function (24) to the entropy. Whereas Boltzmann’s principle asserts that the greatest number of complexions, that correspond to a single macroscopic state, possesses the greatest entropy, (32) affirms that greatest number of digits needed to specify a given set,  $\left( \inf \sum_{i=1}^N q_i \right)^{-1}$ , corresponds to the greatest error in discriminating between two probability distributions. In other words, Boltzmann’s principle is a measure of attenuation, whereas (32) is a measure of accentuation.

This is related to the problem of “how to keep the forecaster honest” [23]. A forecaster uses an estimated probability distribution  $q$  to determine the outcome of events whose true distribution is  $p$ . His fee for correct prediction  $f(q)$  is to be paid after it is known that the event has occurred. His expected fee is  $\sum_{i=1}^N p_i f(q_i)$ , which, if he is honest, must satisfy

$$\sum_{i=1}^N p_i f(q_i) \leq \sum_{i=1}^N p_i f(p_i).$$

The Shannon error function (24) identifies  $f$  with the logarithm. Then  $\inf \sum_{i=1}^N q_i$  in (32) is that state with the lowest degree of predictability.

If  $q$  is the uniform distribution then the error (28) reduces to the differences in entropies so that the exponential of this difference is equal to the mean:

$$\mathfrak{M}_\tau(1/Np) = D^{S_R(p) - S_H(N)}. \quad (33)$$

The closer  $p$  is to the uniform distribution, the larger will be the mean value (33). Interpreting  $\mathfrak{M}_\tau$  as a measure of extent, small values of  $\mathfrak{M}_\tau$  imply that the probability measure is concentrated on a set of small  $q$ -measure for  $0 < \tau \leq 1$ .

As (33) shows, the difference between the Hartley and Rényi entropies is an exponential measure of the magnitude of  $\mathfrak{M}_\tau$ , and hence to the extent of the distribution. The fact that a small value of  $\mathfrak{M}_\tau$  implies a small probability measure requires  $\tau$  to be confined to the interval  $(0, 1]$  [10] is thus related to the condition of concavity of the Rényi entropy. Only for  $\tau \in (0, 1]$  will the exponent in (33) be a true difference in entropies.

- 
- 1961) pp. 547-561.
- [2] C. Tsallis, *J. Stat. Phys.* **52** (1988) 479.
  - [3] A. Feinstein, *Foundations of information theory* (McGraw-Hill, New York, 1958) §2.3.
  - [4] G. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed. (Cambridge U. P., Cambridge, 1952) Chap. III.
  - [5] A. Kolmogorov, *Atti R. Acad. Naz. Lincei* **12** (1930) 388.
  - [6] M. Nagumo, *Japan J. Math.* **7** (1930) 71.
  - [7] J. Aczél, *Functional equations and their applications* (Academic Press, New York, 1966) p. 153.
  - [8] L. L. Campbell, *Inform. Control* **8** (1965) 423.
  - [9] A. I. Khinchin, *Mathematical foundations of information theory* (Dover, New York, 1957) p. 91.
  - [10] L. L. Campbell, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **5** (1965) 217.
  - [11] J. Aczél and Z. Daróczy, *On Measures of information and their characterization* (Academic Press, New York, 1975) p. 186.
  - [12] B. H. Lavenda, “Information and coding discrimination of pseudo-additive entropies”, cond-mat/0403591.
  - [13] H. Havrda and J. Charvát, *Kybernetika (Prague)* **3** (1970), 30.
  - [14] Z. Daróczy, *Inform. & Contr.* **16** (1970) 36.
  - [15] B. L. Burrows and R. F. Talbot, *Int. J. Math. Educ. Sci. Technol.* **17** (1986) 273.
  - [16] E. F. Beckenbach and R. Bellman, *Inequalities* (Springer, Berlin, 1961) p. 23.
  - [17] L. L. Campbell, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **8** (1966) 113.
  - [18] C. Beck and F. Schlögl, *Thermodynamics of chaotic systems* (Cambridge U. P., Cambridge, 1993) p. 53.
  - [19] C. Tsallis, *Brazilian J. Phys.* **29** (1999) 1.
  - [20] A. Rényi, *Probability Theory* (North-Holland, Amsterdam, 1970) p. 562.
  - [21] S. Kullback, *Information theory and statistics* (Wiley, New York, 1959) p. 7.
  - [22] D. F. Kerridge, *J. Roy. Statist. Soc. Ser. B* **23** (1961) 184.
  - [23] I. J. Good, *Uncertainty and Business Decisions*, 2nd ed. (Liverpool U. P., Liverpool, 1957).

---

[1] A. Rényi, in *Proc. 4<sup>th</sup> Berkeley Symp. Math. Statist. Probability, 1960* Vol. 1 (U. California Press, Berkeley,